To see or not to see: Identifying and assessing plagiarism in non-native students’ academic writing without using text-matching software

Katalin Doró
University of Szeged

Maintaining and promoting academic integrity is of great concern in higher education. While some forms of textual borrowing are easily detectable even without text-matching software packages, assignments and manuscripts with an unacceptable level of copying do slip through at various levels of academic life. This paper looks at the markers of plagiarized texts identified by instructors at six higher education institutes in Central and Eastern Europe. It investigates what they perceive to be the reasons behind and indicators of plagiarism and how they indentify textual borrowing in a student text given to them for evaluation. The results suggest that while instructors are well aware of the most common signs of patchwritten student papers, they often do not recognize or consider a piece of writing plagiarized and give markedly different evaluations for it. Possible reasons for this are discussed along with the implications of the results for academic writing instruction, evaluation and policy making.

Key words: plagiarism, patchwriting, academic writing, higher education, text matching software

Introduction

Academic integrity is a widespread concern in higher education institutions (HEIs) around the world. One common form of academic misconduct is the reliance on source texts to the extent that it violates academic integrity policies. Although plagiarism is a widely used term, there is no commonly accepted definition or set of steps that would apply to all HEIs (Adam, Anderson, & Spronken-Smith, 2016, Bretag, 2013; Ho, 2014). There are considerable differences not only across countries and HEIs, but also in how individual students and instructors make sense of policies, how they understand and use the key terms, and what they do to avoid and deter extensive textual borrowing practices (e.g., Foltýnek, Rubička, & Demoliou, 2014; Glendinning, 2016; Horváth, 2014). Research has shown that there is a wide discrepancy between the expectations posed by policies and instructors and the way students understand what constitutes plagiarism (Adam, Anderson, & Spronken-Smith, 2016; Glendinning, 2016; Kokkinaki, Demoliou, & Iakovidou, 2015; Mahmud & Bretag, 2014). Many students believe that if they put effort into restructuring and partly paraphrasing their sources, they violate no rule and do not recognize that they can plagiarize inadvertently.

Up until recently the majority of published studies on academic honesty and textual borrowing were based in North America and Asia. In the last five years, however, there has been a visible increase in the number of articles and books reporting on a wider variety of countries (e.g., Bretag, 2016; Velliaris, 2017). The European...
discussion on the issues has been aided by conferences such as the one organized biannually in Brno, *Plagiarism across Europe and beyond* (for the proceedings see http://plagiarism.pefka.mendelu.cz) and the EU-wide research project called the *Impact of policies for plagiarism in higher education across Europe* (IPPHEAE, see Glendinning 2016; Foltýnek & Glendinning, 2015; Foltýnek, Rubička & Demoliou, 2014). The IPPHEAE project found considerable differences between the participating countries concerning knowledge about policies, integrity measures, training about plagiarism, perceived rate of plagiarism and participants’ experiences about and understanding of plagiarism scenarios. Foltýnek and Glendinning (2015) report that concerning training, Hungary, Luxembourg and Sweden placed on top with all participants indicating training received, while Bulgaria, the Czech Republic and Poland placed at the bottom. The number of students who have committed plagiarism varies between 65% (Lithuania and Italy) and zero (Denmark and Luxembourg). Interestingly enough, the order of countries changes again when the participants are asked about plagiarism committed by their peers. The list is led by Luxembourg and Hungary (with a 100% agreement rate) and closed by Finland and Sweden (with a 20 and 15% agreement rate).

Discrepancies between students’ and teachers’ understanding and perceptions regarding plagiarism have also been found. While studies report that students have a general understanding that plagiarism is bad and should be avoided, they may not recognize what constitutes plagiarism and why referencing texts and ideas is important or they may treat citation as a mechanical step that helps them avoid plagiarism charges (Bretag, 2013; Chanock, 2008; Glendinning, 2016; Pecorari, 2003). It is also worth noting that the IPPHEAE data reveal significant differences between students’ and instructors’ perception on why academic writing is difficult. While teachers consider paraphrasing, citation and formatting references to be the leading difficulties, students give much less importance to these aspects, and believe that finding good sources is the main problem generator (Foltýnek, Rubička & Demoliou, 2014). Both groups reported the ease of copying and pasting from the Internet as the main reason for plagiarism. Students placed lack of time and difficulty coping with coursework as the second and third reasons. In contrast, instructors pointed out that plagiarism is not seen by students as morally wrong, lecturers do not care and students lack adequate reading skills. These and similar differences are important to be aware of when giving assignments and designing academic writing assistance.

Adam et al. (2016), similarly to Kaposi and Dell (2012), review three distinct views on plagiarism, namely moral, regulatory and academic writing discourses. The first one discusses the ethical, dishonesty and fairness concept of plagiarism and treats it as a common form of academic misconduct and intellectual theft. The second one focuses on policies concerning plagiarism, adherence to rules and plagiarism detection. The third one stresses the need to educate rather than punish and to teach students academic writing skills. The moral view can often be heard both from students and
instructors. Adam et al. (2016) point out that although it is very common to treat plagiarism as a form of misconduct, it is not very helpful in supporting students in the academic writing process or in their better understanding of expectations.

Supporters of the educational view on plagiarism focus on the development of students as academic writers and almost exclusively discuss unintentional plagiarism. In opposition to plagiarism, terms such as *textual borrowing* (Keck, 2014; Petrić, 2012; Shi, 2004) and *patchwriting* (Pecorari 2003; Howard, 1993, 1999; Howard, Serviss & Rodrigue, 2010; Li & Casanave, 2012) have been constructed. Howard defines patchwriting as “copying from a source text and then deleting some words, altering grammatical structures, or plugging in one-for-one synonym-substitutes” (Howard, 1993, p. 233). Many researchers and academic writing instructors support the view that students, especially if writing in an L2, are emergent writers who are balancing between many factors, such as integrity rules, perceived expectations, time-constraints, difficulty of the assignment and literacy skills (e.g., Adam et al., 2016; Harwood & Petrić, 2012; Schembri, 2009; Shi, 2012). They treat patchwriting and textual borrowing as a step in the learning process that most students go through and grow out of if adequate writing support is provided. This learning to write view is successful, though, only if textual borrowing is identified at an early stage of writing and the student is followed through the draft stages with multiple rounds of feedback. Oftentimes, however, an assignment is assessed only as a product and not as a process.

By using text-matching tools researchers have pointed out surprising degrees of plagiarism or other forms of source misuse in student writing even in countries where strong integrity policies are in use. Vieyra, Strickland and Timmerman (2013) found that 28% of the investigated research proposals written by US graduate and postgraduate science and engineering master’s and doctoral students contained some textual borrowing without appropriate citation. The authors revealed that 68% of these were direct copies, 17% contained some lexical and 12% some grammatical changes. Only 7% of the plagiarized sentences were given an appropriate citation and 80% of them had no citation at all. Jamieson and Howard (2013), while investigating research papers written by first-year American students, found that only 6% contained summaries, 32% paraphrases of sources, 16% patchwriting and 56% direct copies from sources, most of which were non-academic, including information leaflets, instructions, blogs and Wikipedia. The vast majority of students worked from a few selected sentences of their sources that came from the first one or two pages. The same sentence-level textual reuse called *sentence-mining* was also documented in another study (Howard, Serviss & Rodrigue, 2010). Moore (2014) found that 31% of her reviewed 91 Finnish theses contain one or more of the following problems: constant inaccuracy, referencing problems and plagiarism.

The identification of plagiarism is also more difficult than it seems. Text-matching software packages aid instructors at many universities nowadays and their use has been widely discussed in the literature. Many HEIs adopt Turnitin and other similar
software packages not only to identify the degree of borrowed textual chunks, but also to use them as learning tools through which students themselves can check their writing before submission. Research has found this practice welcome and positively evaluated by both students and instructors (e.g., Bailey & Challen, 2015; Buckley & Cowap, 2013; Graham-Matheson & Starr, 2013; Kostka & Maliborska, 2016). Much less is known, however, about what happens on a daily basis with assignments that are not checked for text matching either because they are classified as smaller, less important assignments, because the instructor is overwhelmed or because text-matching tools are not available. There is much anecdotal evidence that instructors check texts in some way only if obvious change in writing style is detected (except for the HEIs where all assignments are checked centrally for plagiarism and assignment recycling), but not all textual reuse can be identified from stylistic or grammar clues. In the absence of software, instructors may plug anomalous looking sentences into Google which will hopefully return the plagiarized textual chunk. This method is fast and effective; however, what can be googled is limited in length and would not give results for off-line sources and assignments downloaded from paper mills or written by ghost writers. Research has clearly pointed out how variable the procedures for dealing with academic misconduct are among HEIs and even within the same institutes. Some instructors are inclined to let low-level cases slide even if plagiarism is suspected in order to avoid personal conflicts, to balance for unclear integrity rules and procedures and to save the hassle that goes with a plagiarism case.

The current study is intended to fill the gap in the research literature by examining how instructors working with students of English in Central and Eastern European universities view the reasons behind plagiarism, how they assess a plagiarized student paper and whether they are able to identify textual borrowing. Limited published data are available on plagiarism and patchwriting issues from this area of Europe and likewise very little is known from other regions about instructors’ ability to identify plagiarized texts and their assessment of such texts in situations where text-matching tools are not available (see Doró, in press; Foltýnek, Kravjar & Glendinning, 2014; Glendinning, 2013; Horváth, 2012; Pecorari & Shaw, 2010). Although exploratory in size and nature, this study is hoped to contribute to the growing awareness of the problem of plagiarism and provide insight into the daily assessment of students’ academic writing.

Methods

Twenty instructors working in six departments in Hungary (n=14), Romania (n=2), Poland (n=2) and Croatia (n=2) participated in the study upon a personal invitation by the author of this study. In terms of their biographical and professional background, they show considerable variability. Three of them speak English, two Croatian, one Finnish, three Polish, two Romanian and nine Hungarian as their mother tongue. Except
for two, all of them have more than five years of higher education teaching experience, and fifteen of them have been teaching, evaluating student papers and doing reviews of academic manuscripts written in English for more than 10 years. All the six departments work with students of English who are enrolled in programs to become English philologists, researchers, interpreters, translators or teachers. Assessment of assignments written in English by non-native English speaking students is a daily task for the participants. Student assignments are handed in at these departments both electronically and in printed form, depending on the request of the instructors or the regulations for final theses; therefore, it is still common to read printed assignments and mark them by hand.

The participants were given an envelope which contained written instructions, an extract from a student paper and a sealed questionnaire in English. As the main aim was to see whether colleagues were able to detect the main problematic points of the sample text, the data collection was first masked as an error correction project. Participants were asked to assess the sample page from a student’s final paper (it was not revealed to them that it was an extract from a failed MA thesis) the way they would for any other assignment. The preceding two pages were also provided for context but were not assessed. After the first round of reading and marking, they were instructed to open the questionnaire which included questions related to some bio data of the participants, their degree of familiarity with the topic of the marked text and the types of errors noticed. They were then informed that the text was written by a student who had been caught plagiarizing in previous assignments. They were asked to go back to the text and, using a different color, mark anything new they noticed, underline sections they believed were plagiarized and put a question mark if plagiarism was suspected but they were unsure about it. The following seven open-ended questions then related to their assessment of plagiarism and perceived reasons and signs of plagiarism. These were included to see whether the participants have overlapping general beliefs and follow similar practices and whether these match their assessment of the specific extract.

Results and discussion

Plagiarism: signs and possible reasons

As for the signs of plagiarism in student texts, the following were mentioned: an obvious change in style, vocabulary use not typical of the local student populations, unfinished sentences, formatting problems (e.g., uneven font and size), references out of reach for students, inconsistent referencing, cited publications not in the reference list, major claims and data without references, reading the same passage in two papers and passages that look familiar to the instructor. Some of these signs indicate careless copy-paste plagiarism in which students do not pay attention to final formatting and, therefore, the lack of originality is easily detectable. Changes in style and grammar may
flag potential cases of plagiarism for closer inspection. Other cases (such as referencing anomalies or language use not typical of the local L2 students) need further investigation and may not even be easily spotted, especially if a large number of assignments and pages need to be read. These signs are not universal and may change depending on the assignment, field of study, students’ proficiency level and seniority and whether the students are writing in an L1 or an L2, but largely overlap with those indicated in other research and those discussed by colleagues during their daily work (Doró, in press). It is also important to note here that some students who use excessive textual borrowing as a regular, deliberate plagiarism strategy may not show any of these signs and patch together their texts so that they look fine on the surface (Doró, in press).

The reasons for plagiarizing indicated by the participants can be grouped under the following 13 categories, reworded by the author (the number of occurrences is indicated in parenthesis):

1) Inadequate language and general writing skills (12)
2) Economy of effort, perceived ease of cheating (10)
3) Lack of information about what constitutes plagiarism (8)
4) Lack of citing and paraphrasing skills (8)
5) Demanding schedules, lack of time, last minute work (6)
6) Getting away with it (beating the system, tricking the instructor, circumventing policies) (5)
7) Lack of self-confidence: not as smart as the author, unable to paraphrase the source to indicate the original ideas (5)
8) Pressure to perform well, good grades (5)
9) Desire to look smarter (4)
10) Task too demanding (3)
11) Lack of ideas (3)
12) Permissive plagiarism practice (2)
13) Plagiarism culture brought from secondary school (2)

Participants usually indicated more than one reason for student plagiarism. The top reasons refer to the lack of literacy skills and information about academic writing conventions, the ease of cheating, and little effort put into producing original work. Not surprisingly, these largely overlap with reasons indicated by instructors in previous studies (e.g., Foltýnek, Rubička, & Demoliou, 2014). However, general language skills and economy of effort are more prominent in our study. It is interesting to note that none of the participants used the words ‘lazy’ or ‘laziness’ when explaining the small amount of time or effort put into producing original work. Students, on the contrary, tend to label this practice with a general comment of laziness (Dorô, 2014; Horváth, 2014; Horváth & Reif, 2010). The mere testing of the system, the thrill to trick instructors and lack of ideas as reasons for dishonest work have not typically been
documented in other research, but, again, were voiced by local students (Doró, 2014). Demanding tasks and permissive local practice were rarely mentioned, possibly because these are not perceived as reasons per se or would count as self-criticism. The lack of interest in the subject and uninteresting tasks reported in a Swedish study (Razera, 2012) were completely missing from the perceived reasons.

The division between imposing realistic and unrealistic expectations is very thin and often unclear. An assignment can be fair and realistic, but impossible to complete honestly for a student who has lower than expected language skills and subject knowledge and/or leaves the research and writing to the last minute. In this case fear of not completing the assignment may easily override intelligence and critical judgment of what is honest work. Even good students may panic at the last minute, or not invest proper time and effort in the course and hope to get away with it. It is also important to note, as indicated by two participants, that students may bring with them plagiarism from earlier school contexts. This practice has been noted in the literature, calling attention to the fact that it is difficult to teach students about the inappropriate nature of some forms of source misuse if academic dishonesty has been practiced in elementary and secondary schools (Anderman & Midgley, 2004; Jensen, Arnett, & Cauffman, 2002, McCabe, Butterfly, & Trevino, 2012; Velliaris, 2017).

Plagiarism assessment

The extract from the student text was marked for grammatical, sentence structure and vocabulary choice errors, typos and lack of coherence during the first reading. The participants used various marking systems, all of which clearly indicated the type of errors found. During this first marking stage, only two colleagues indicated problems with source use; one of them spotted a missing item from the reference list and the other one indicated one sentence as copied from a source. Only this person answered yes to whether they had found instances of plagiarism in the text.

When asked to review the extract again, three more colleagues identified at least one sentence that was copied verbatim, but only two of these five participants found all the problematic sections and marked the sources (and indicated having used Google search for this). Of the remaining fifteen participants, seven put a question mark next to one or two sentences, indicating a suspected textual borrowing case, but probably did not verify their suspicion. This means that 40% of the participants did not identify a textual borrowing problem even after having learned about the student’s previous plagiarism problem and wrote that they did not see signs of plagiarism in the text. Even though everyone was able to list signs of plagiarism, they did not necessarily spot these in a text where these signs were all present on a single, selected page.

As for assessment, the participants had three subtasks, namely to indicate the type of feedback or evaluation they would give this paper if it was either a seminar paper, a thesis draft or a thesis. The seminar paper received 3 low grades (not specified), 6
failing grades, 8 failing grades with the option to rewrite and 3 rewrite without a grade assessments. Some written comments such as “horrible paper” or “weak paper” were also given. Eight of the participants would request in-person discussion with the student. The thesis draft was marked to be rewritten by all the participants with an additional failing grade by three of them. The same eight participants indicated willingness to talk to the student. If part of a thesis, the extract would receive a failing grade from 18 participants and a low grade by 2 of them. As it was pointed out above, most of the participants did spot some serious problems with the texts, but did not necessarily put together all the signs of plagiarism; therefore, their assessment of the text did not necessarily reflect the assessment of an assignment they would classify as plagiarized. Some of the participants seem to be strongly oriented toward approaching L2 writing problems as a learning issue and would give students the option to rewrite their texts, especially if the paper is at an early stage and would also talk to the student. The alarming issue is that even if the extract would be given back to the student for improvement, in the majority of cases it was not marked for inappropriate textual borrowing, but rather for punctuation, subject-verb agreement, articles, vocabulary choice, sentence structure, coherence and formatting. This would give the impression to the student that source misuse was acceptable or, at least, not identified, therefore not an issue to worry about. This makes it difficult to properly deal with plagiarism and provide appropriate writing support to students (Doró, manuscript).

Diversity in the assessment of student texts has also been documented by Pecorari and Shaw (2010) who interviewed ten instructors about their decision as to appropriate source use or plagiarism regarding five student paper extracts. The participants did not only disagree with others on their evaluations, but they were often uncertain about their own assessment of a text. They differed in their labeling, in their level of confidence and in the spontaneity of their answers. Some of the borderline cases were also given back to students for improvement in this study. The authors point out the alarming nature of this disagreement among staff members about what constitutes good, appropriate and acceptable writing and the inconsistent feedback they provide to students based on their decisions. Even text-matching tools do not decide for the instructors, but only flag closely matching texts. Instructors need to develop a more homogeneous view of what is acceptable, at least in a given higher education context.

Conclusion

This paper reported on an experimental study conducted among instructors at Central and Eastern European HEIs to see how they view and assess plagiarized student texts. While the participants in this study seem to be well aware of the most common signs of patchwritten student papers, very few of them recognized heavy textual borrowing and they gave markedly different evaluations for the same student text. Knowing about the student’s history of plagiarism did not have a substantial effect on the correction in
some cases. Plagiarism was suspected by a larger number of participants during the second reading, but rarely checked. This study also found that views about and assessment of plagiarism and textual borrowing are diverse even among colleagues working in the same or similar higher education settings.

The above reviewed literature usually looks at one specific segment of the plagiarism/text reuse issue. Bretag (2013) urges a holistic view on plagiarism that includes providing information and ongoing training to both students and staff, carefully designed policies and their implementation, better and more consistent understanding of terminology, support embedded in all courses, and finally detection used both as a learning tool to avoid plagiarism and a reliable and systematically applied way to flag matching text portions. This study also shows that many universities are far from a holistic view. However, a better general understanding of the reasons behind plagiarism and a more consistent assessment of student texts would lead to a greater awareness of the issue. Until heavy textual borrowing is not detected in student writing, instructors will hardly understand the dimension of the problem and the need for a prompt intervention. Especially where the use of a text-matching tool is not part of the general assessment practice and source reuse is not easily flagged, many student papers slip through and students may be given the wrong message about the (non)acceptance of academic misconduct. This makes subsequent academic writing assistance less beneficial and the introduction or the use of academic integrity rules problematic. Awareness of the diversity of views among staff and students on plagiarism and of the academic writing problems that probably have much larger dimensions than what is currently believed would serve as a good starting point for handling source misuse at HEIs.

References


Doró, K. (manuscript). From phrase to discourse level patchwriting: Is it possible to unlearn?

Foltýnek, T., & Glendinning, I. (2015). Impact of policies for plagiarism in higher education across Europe: Results of the project. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis, 63*(1), 207–216.


Horváth, J. (2012). „A szakmai közösségnek tudnia kell ezekről a problémákról”: A plágium kezelése a magyar egyetemi gyakorlatban [“The academic community should know about this: Dealing with plagiarism at Hungarian universities”]. *Iskolakultúra, 22*(7-8), 96–110.


